SU+ @ Strathmore
University Library

Electronic Theses and Dissertations

2022

A Systematic comparison of performance of Ridge, Lasso, Elastic net and Relaxed Elastic net when fitting high dimensional data for sales prediction.

Muoki, Monica Mueni

Strathmore Institute of Mathematical Sciences
Strathmore University

Recommended Citation

Muoki, M. M. (2022). *A Systematic comparison of performance of Ridge, Lasso, Elastic net and Relaxed Elastic net when fitting high dimensional data for sales prediction* [Strathmore University]. http://hdl.handle.net/11071/13176

Follow this and additional works at: http://hdl.handle.net/11071/13176

This work is availed for free and open access by Strathmore University Library.
It has been accepted for digital distribution by an authorized administrator of SU+ @Strathmore University.
For more information, please contact library@strathmore.edu

A Systematic Comparison of Performance of Ridge, Lasso, Elastic Net and Relaxed Elastic Net when Fitting High Dimensional Data for Sales Prediction

Muoki Monica Mueni

Submitted in partial fulfilment of the requirements for the degree of Master of Science in Statistical Science at Strathmore University

Institute of Mathematical Sciences
Strathmore University
Nairobi, Kenya

October 2022

This thesis is available for Library use through open access on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Declaration

I declare that this work has not been previously submitted and approved for award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

© No part of this thesis may be reproduced without the permission of the author and Strathmore University.

Name: .... **Muoki Monica Mueni** ....

Signature: ....

Date: .... September 8, 2022 ....

Approval

The thesis of (Muoki Monica Mueni) was reviewed and approved by the following:

Dr.Collins Odhiambo
Supervisor,
Institute of Mathematical Sciences,
Strathmore University.

Dr. Godfrey Madigu
Dean,
Institute of Mathematical Sciences,
Strathmore University.

Dr. Bernard Shibwabo
Director,
Office of Graduate Studies,
Strathmore University.

Abstract

Forecasting or prediction is one of the most crucial aspects of planning for many companies.

Data-driven decisions can only be as accurate as the prediction they are based on. Some of the decisions include production planning, inventory management, and various resource allocation. Sales information is really multi-dimensional, and as a result not easy to analyse. Our motivation is to reduce the high dimension of this information, select optimal contributing variables with the aim of making accurate and reliable sales predictions. The purpose of this study is to compare the performance of four restricted regressions. This involves looking at Ridge, Lasso, Relaxed net and Elastic net regressions and assessing their performance in prediction when dealing with high dimensional data. The proposed method will involve comparison of the four mentioned regularized techniques, citing their restrictions and evaluating their prediction model performance. We will also involve data simulation to test the different models. The simulations are done under different scenarios to present the reality of a market setting. Afterwards, we will select the best model and use it to fit our real sales dataset provided by one of the leading ECMCs in Kenya. On this basis, elastic net offered best predictions based. The evaluating metrics for this models are Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and R-Squared (R^2). However, the desired model based on R^2 kept shifting under different scenarios to Lasso, Ridge and Elastic net.

The results indicated that the regularized approaches especially elastic net are capable of dealing with non-linearity and fluctuating dynamics in manufacturing industry while predicting electrical cable sales accurately.

Key words: Prediction, ECMCs, Regularized techniques, Simulation, Sales data, RMSE.

Table of contents

UT OMNES VNVM SINT

VT OMNES UNVM SINT

List of figures

List of tables

List of abbreviations

Lasso	least absolute shrinkage and selection operator	LR	lasso Regression
RR	Ridge Regression	ENR	Elastic Net Regression
ECMC	Electrical Cable Manufacturing Cable	RMP	Raw Material Planning
RMSE	Root Mean Squared Error	MAE	Mean Absolute Error
R^2	R-Squared	PM	Prediction model

Acknowledgement

I would like to start by thanking my supervisor Dr. Collins Odhiambo, for all the support, time and effort he has put into helping me see this thesis through to completion. I would like also to thank my parents, siblings and friends for their unwavering support and help throughout the years. Finally, I would like to thank all the faculty, staff and students in our department for making my time here at Strathmore University such a wonderful experience.

Dedication

This thesis is dedicated to God Almighty for giving me wisdom, knowledge and good health.

Chapter 1

Introduction

1.1 Background to the study

Accurate sales prediction is an important aspect for companies to augment their profits, reduce the cost, and achieve greater flexibility to sales parameters. In other words, precise sales prediction is utilized for capturing the trade-off between customer demand satisfaction and inventory costs Gupta et al. (2000). Electrical Cable manufacturing Companies in Kenya are mainly owned and ran by independent Sole-proprietors or group of companies, who buy raw materials for their own accounts from foreign suppliers, manufacture electrical cables and sell to their customers (Traders, Contractors and Developers). These manufacturers sell their finished cables to distributors or users through face-to-face commonly walk in customers to their factories, web ordering, or telemarketing. However, in some cases, these distributors also have their importers and their own imported cable. This brings in the competition and the ever-changing environment in this industry. A basic consumer link from the manufacture to the final user is shown Figure 1.1. This is to illustrate the flow chain from raw material acquisition to manufacturing to the end user.

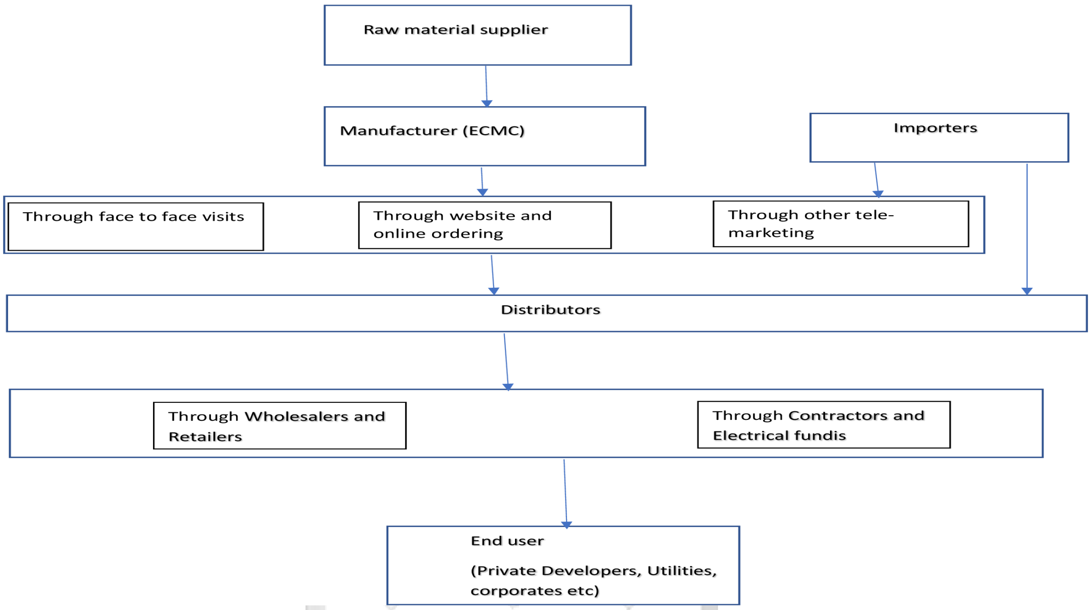

Figure 1.1: ECMC Distribution Channels Presentation From Production to the End User.

Electrical cable manufacturing companies are faced with several challenges. These include: huge amount of raw materials, increased competition, fluctuating price and demand, and the pressure to meet customers' needs by delivering the right quality of cables to the right place and at the right time. To maintain all these successfully, accurate sales prediction is a necessity. In this research, we have collected the required sales data from one of the large ECMCs, which distributes its products widely in Kenya. To retain its leading market position, this company needs to maintain large inventories to avoid any chance of stock out in order to meet its customers' demand. Therefore, this company like any other ECMC in the industry, keeps inventories covering needs of the next 1-2 months in advance. To maintain this inventory control, transportation logistics and implicated financial costs contribute a high percentage of total expenses in ECMCs. For instance, ECMCs buy raw materials from foreign suppliers and pay for them instant on purchase. On the contrary, they sell their products and gradually receive the related money returns (sales on credit) which can last to even 90 days. As a result, there is an undesired expenses gap which calls for accurate and monthly sales prediction to minimize or even eliminate this gap. Therefore, to help reduce undesired inventory costs, achieve increased profits and maintain customer-satisfaction, these manufacturers require or need an accurate sales prediction to make informed decisions hence achieve balance and stability.

1.2 Motivation of the study

Sales is the backbone or pillar of any manufacturing/retail business (Hanssens et al., 2003) and (Reddy et al., 1994). As simple as it is to use prior sales to predict the future, sales data pose challenges due to the fluctuating market environment and the many attributes (P) affecting sales volumes. However, it is unlikely that all P attributes are important when the number of P is extremely large. A model that has the capability to best explain and give the desired relationship with the least number of attributes, is our desired choice. As such, a common approach would be to identify a smaller subset of predictors that have the largest impact on our response variable or employ approaches that tend to penalize the remaining predictors such that their estimated coefficients are either small or even zero. This problem of fitting an appropriate model in the high-dimensional setting presented in this paper call for techniques that can tackle these forms of problems which is our main focus in this research. Therefore, we employ penalized regression models as our main focus to enable us;

1. Reduce the high dimensionality of sales information.

2. Select optimal important attributes affecting sales.

3. Build an accurate prediction model for future sales volumes.

1.3 Problem Statement

When the number of features in a dataset exceeds the number of observations, we can never have a deterministic answer. Therefore, it becomes impossible to find a model that can describe the relationship between the predictor variable and the response variable since the observations are not enough to train the model. This study wishes to systematically compare, citing all restrictions and limitations of regularized approaches to come up with an accurate and clear model to predict sales. Regularized or penalized approaches impose penalty to a model for having too many variables. As a result, this process selects the most influential predictors.

1.4 Objectives of the study

To build accurate sales prediction models, we will try to evaluate the following objectives;

1. Evaluating the performance of penalized techniques in analyzing high dimensional data to build prediction models.
2. Systematically comparing the restrictions of penalized approaches to come up with an approach which performs better than the rest.
3. To select/choose the attributes that contribute most to our model accuracy.

1.5 Significance of the study

Sales prediction is concerned with estimating future sales of companies. It plays a vital planning role in ECMCs and influences many unavoidable decisions which can only be as accurate as the predictions they are based on. This research looks at cable sales prediction in the cable manufacturing industry and seeks to explore regularized approaches when dealing with sales data to establish an approach that accurately simplifies sales information giving optimal variable section and an accurate prediction model, for data informed decision making (Mentzer and Bienstock, 1998a) and (Mentzer and Bienstock, 1998b). Some of these informed decisions would include;

1. Raw material planning (RMP)
2. Inventory management
3. Labour-force scheduling
4. Capital budgeting
5. Distribution planning

With these done accurately, it is possible for a business to thrive, despite the competitive and volatile business dynamics.

Chapter 2

Literature review

2.1 Introduction

For every product category or brand, there exists a sales generating process (Hanssens et al., 2003). The current business environment is highly competitive and constantly changing. Accurate estimation of future sales, also known as sales prediction or sales forecasting, whether short or long-term, can offer fundamental information to businesses and mostly to companies in the manufacturing industry. In this section, we present reviewed related works by several authors on sales prediction.

Most approaches that have been used in sales prediction have normally been time series forecasting methods. These methods can either be defined as linear or non-linear given to the nature of the model in question (Doganis et al., 2006). Linear models, such as autoregressive moving average (ARMA), and autoregressive integrated moving average (ARIMA), (Song and Esogbue, 2006) predict future sales based on previous sales only. For instance, in a market setting, sales are more influenced by exogenous variables such as competition , climatic data, effect of media, price changes or nonlinearity (Wang et al., 2013). Their prediction ability is limited by their assumption of a linear behavior and as a result, it is not always satisfactorily accurate (Zhang, 2003).

To address possible non-linearities in forecasting, a number of non linear methodologies were introduced. These included the non-linear ARMA time series models. These had a main limitation: which is the type of non-linearity is not known in advance and hence the researcher needs to select the structure of the model by trial and error. Advanced artificial intelligence techniques, such as artificial neural networks (ANN) (Kubat, 1999) and fuzzy logic systems use more complicated generic model structures that can tackle complex data and offer accurate time series models, by eliminating the time consuming trial and error procedure.

The application of standard regression models to a group of explanatory variables to make a prediction, sometimes leads to overfitting (Ranstam and Cook, 2018). This is both in terms of the number of variables included in the final model and an overestimation of how the model performs (Ranstam and Cook, 2018). Therefore, a selection of penalized/regularization regression techniques are applied and these include; lasso, ridge, elastic net and relaxed lasso (Ranstam and Cook, 2018).

Several studies have been done to compare these techniques. For instance, found that the adaptive lasso, adaptive elastic net, elastic net, and lasso all had somewhat equal accuracies and they all outperformed ridge regressions in terms of RMSE and correlation (Ogutu et al., 2012).

With the advancement of computers, ANN have been extensively applied reason because they have considerably promising prediction performance (Frank et al., 2003) and (Chang and Wang, 2006). Artificial neural networks popularity is attributed by their capability to simulate a wide range of underlying non-linear behaviours and to cover a broad variety of fields (Chu and Zhang, 2003). They are better than many traditional or conventional statistical prediction approaches. However, their sales prediction methods use gradient-based learning algorithms, such as the back-propagation neural network(BPNN), which are prone to over-fitting problems and long computation time still arise Sun et al. (2008).

Due to the increase in growth of use of ANNs, numerous studies have tried to implement neural networks to time series forecasting and compared them with conventional methods. However, the outcomes are not distinctly in favour of one specific methodology (Kuo et al., 2002) and (Doganis et al., 2006). Looking at the various comparative studies between traditional models and neural networks which have been done, different outcomes have been obtained with regard to whether ANNs are better than the linear methods in forecasting or not (Adya and Collopy, 1998) and Zhang et al. (1998).

In this regard, many researchers have provided empirical confirmation on the comparative preference of one method over the other under different forecasting scenarios (Nelson et al., 1994), (Paliwal and Kumar, 2009), and (Kuo et al., 2002). The majority of these studies, illustrate varied results on the effectiveness of the ANNs compared with the traditional models like ARIMA (Wang et al., 2013). Since invention of neural networks, only few studies have concluded that traditional methods have superior performance to or at least the same performance as neural networks. A comparison study on neural networks' performance to traditional methods in time series forecasting (Tang et al., 1991) demonstrated that for time series with long memory, both ANNs and Box-Jenkins models had the same results. Additionally, (Foster et al., 1992) illustrated that exponential smoothing is better than neural networks in yearly data forecasting.

Although it appears that neural networks may outperform conventional statistical methods in forecasting time series with trend and seasonal patterns, (Nelson et al., 1994) found that neural networks cannot properly model the seasonal patterns in their data. Also, (Callen et al., 1996) showed that forecasting performance of linear time series models was better than that of neural networks even if the data were nonlinear. (Church and Curram, 1996) and (Ntungo and Boyd, 1998) demonstrated that neural networks performed nearly the same as econometric and ARIMA approaches. Afterwards, (Heravi et al., 2004) showed that linear models generate more reliable prediction results than neural networks for European industrial production prediction.

VT OMNES VNVM SINT

(Ainscough and Aronson, 1999) compared neural network and regression analysis, in modelling and predicting the results of retailer activity on the sales of definite products applying scanner data. The results of their study, illustrated that neural networks had better performance than regression model. In addition, (Qi, 2001) reported that ANNs are likely to do better than other methods when the required data are saved as recent as possible. Although statistical techniques have been proven effective for a long time, they still have definite drawbacks (Chen and Ou, 2008), (Yip et al., 1997) and (Kuo and Cohen, 1998). For instance, when the data is influenced by particular conditions, like promotion, the prediction results of conventional methods are undesirable (Kuo and Cohen, 1998), (Kuo et al., 2002). As such, it is logical to anticipate having some level of non-linearity in the sales setting. However, traditional methods are not capable of modelling non linear relationships, hence most researchers apply neural network models to curb these problems.

The contribution of neural networks compared to the traditional methods seems degraded in some cases (Hill et al., 1994), their capability to model nonlinear relations and their adaptation are highly attractive for most forecasting subjects (Zhang, 2004). Therefore, advanced methods like neural networks could serve as more appropriate approximator (Zhang, 2004) and would be more suitable for the time series sales forecasting than linear models like ARIMA.

However, (Adya and Collopy, 1998) reviewed paper presented that regardless of the growing applications of neural networks in prediction and more specifically in business prediction, opinions and outcomes concerning their contribution are varied. They concluded that assessing studies is complicated and lacks an obvious criteria. It means that, although neural networks have acquired growing attention in the forecasting domain, resulting in competent applications in time series sales forecasting (Crone, 2003), no single approach that works best in every condition, and combining diverse methods is an efficient and effective way to advance prediction accuracy (Zhang, 2003). In this regard, recent researches offered good explanations of the hybrid ARIMA-ANN models or combination of other conventional and ANN techniques (Zhang, 2003), (Aburto and Weber, 2007), (Aslanargun et al., 2007), Jain and Kumar (2007) and (Díaz-Robles et al., 2008). An illustration by (Zhang, 2003) recommended a hybrid ARIMA-ANN approach in which ARIMA was applied to model the linear part and ANNs used to model the prediction errors. He showed that the hybrid method outperformed both separate methodologies.

In addition, (Khashei and Bijari, 2010) proposed a new hybrid neural networks model. They applied an artificial neural network (**p,d,q**) model for time series forecasting to have a more precise forecasting model than neural networks . The results of three real data sets demonstrated that the proposed model was an appropriate way to advance forecasting accuracy accomplished by neural networks.

Both theoretical and empirical results have confirmed that combination of various models is a successful way of enhancing the performance of forecasting models Khashei and Bijari (2010). Accordingly, it appears logical to apply hybrid models in most sales predicting domains. According to the presented reviews, numerous prediction methods have been offered and each method has its specific advantages and disadvantages in comparison with other techniques. However, none of the accomplished studies described offered a novel technique for handling the problem of not having enough past records for predicting. However, owing to these specific constraints in a market setting that is having numerous new items with few historical data, existent forecasting methods are generally inappropriate. This motivates us to systematically compare the performance of regularized approaches in cable sales volume prediction.

VT OMNES VNVM SINT

Chapter 3

Methodology

3.1 Introduction and Data Description

Traditionally, constructing predictive models, is usually the case where we have the number of observations greater than predictors ($n > p$). P is the number of features, variables or predictors while n refers to the number of observations. In other words, Variables of interest in an experiment (those that are measured or observed) are called response or dependent variables. Other variables in the experiment that affect the response and can be set or measured by the researcher are called predictors, explanatory or independent variables. As so $n > p$, has to be satisfied in order to obtain ordinary least squares estimates (OLS) for our predictors in a standard linear regression setting (Greene, 1981). However, there scenarios where we have $p > n$, where the number of predictors is larger than the number of sample observations commonly in market setting. Our data for this research (sales data) is drawn from the sales portal of one of the largest ECMCs in Kenya. This portal contains all the inventory costs used in cable production input by a dedicated personnel. Among the many variables listed, we picked up only 20 variables which were of interest to this study. These variables are, raw material fillers used, aluminum metal used, copper metal used, XLPE material used, PVC material used, lubricants costs, packaging material used, drum Material used, SWA material used, scrap material sold, house wire cables produced, auto cables produced, aluminum cables produced, steel conductors produced, battery cables produced, flexible cables produced, armored cables produced, trade, flexible cables produced and sales staff employed. We compiled 5 year monthly data to come up with our real dataset for the study.

Additionally, we set up three possible scenarios in a market setting to test our models as outlined;

1. n > p

2. n = p

3. n < p

More on how this will be done is discussed in the next section. Our methodology will involve, data simulation to train and test our data and the performance of our different models discussed in the next sections. Our data throughout simulations will follow normal distribution. This assumption is as a result of a normal curve laid on our real data as shown in Figure 3.1.

Our aim is to produce a prediction model whose output (sales volumes) is accurate. We will be doing partitioning of train and test data and also choosing a value for the tuning parameter(s)(α and λ).

For this, we will be making use of 5-fold cross validation as the approach of choice to select the best value for our tuning parameters. This will consist of splitting our dataset into 5 equal-sized samples, before collectively fitting the desired model on 4 remaining samples (i.e.,80% of the data is used as the training set) and evaluating the model's performance on the remaining single sample (i.e., 20% of the data is used as the test set). This is then repeated for all five possible cases, where each of the five samples would be used once as the test set. We will therefore compare our models using RMSE, MAE and R^2. The model with the lowest RMSE, MAE and the highest R^2 will be our desired model. We begin by looking at an overview of the regularized approaches presented in our study.

3.2 Regularized Regression Approaches

These are penalized regression methods that allow one to create a linear regression model that is penalized, for having too many variables in the model, by adding a constraint in the

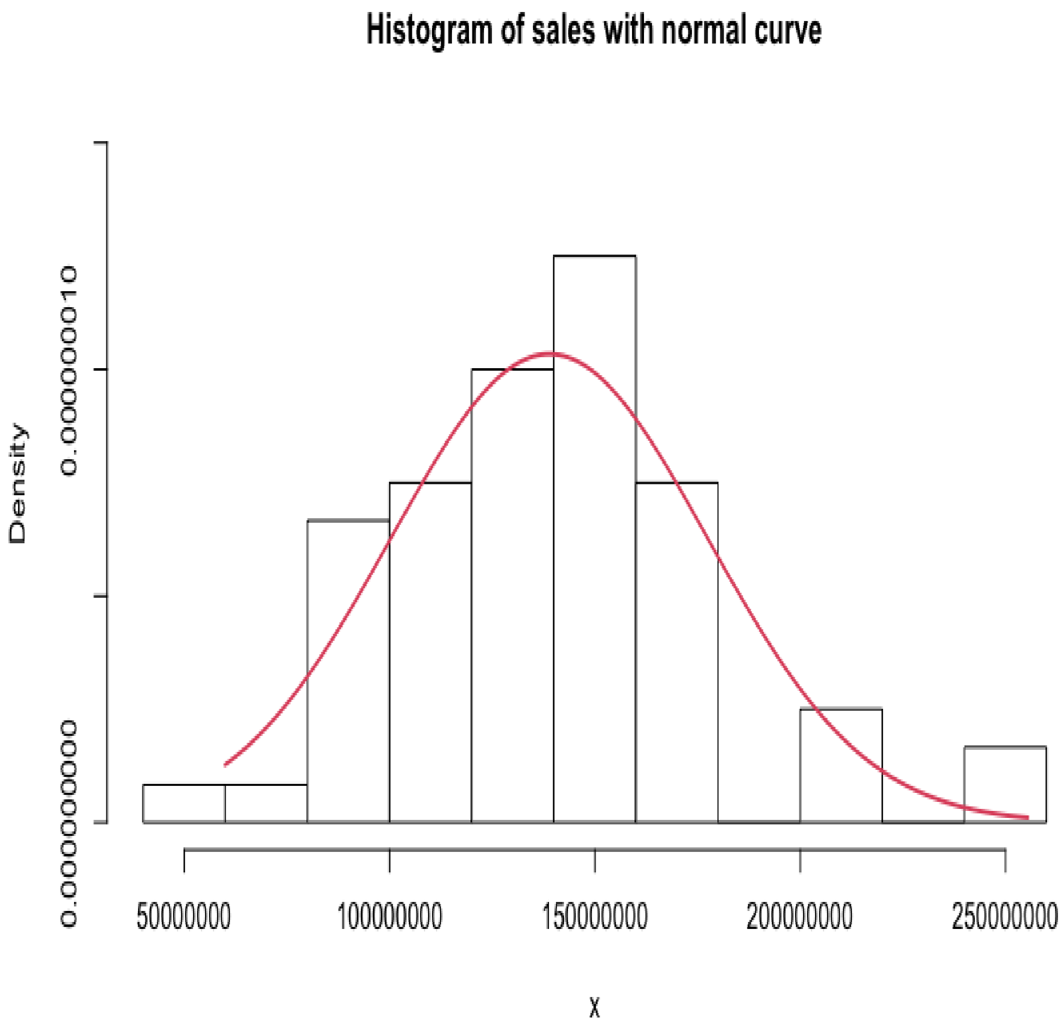

Figure 3.1: Histogram Presentation of Sales Data with a Normal Curve Fit.

equation (James et al. 2014,P. Bruce and Bruce (2017)). Also referred to as shrinkage or regularization methods. These approaches are more continuous, and don't suffer much from high variability thus accurately tackle high dimensionality (Bühlmann and Van De Geer, 2011). They have been highly considered,and thus motivates this research. therefore, we purpose to systematically compare them as outlined,

1. Ridge Regression

2. Lasso Regression

3. Relaxed Net

4. Elastic net Regression

we systematically look at the formation of each, taking note of their restrictions and advantage in prediction.

3.2.1 Ridge Regression (RR)

Ridge regression was initially introduced as a solution directed at addressing the issue of orthogonality and ill poised problem (Bühlmann and Van De Geer, 2011) resulting from multi-collinearity or high-dimensionality. In ridge regression, coefficient estimates are shrunk, reach exactly zero (unless $\lambda = \alpha$), and thus the contrast with the Lasso through the potential towards zero as the penalty parameter λ increases.

However, ridge coefficient estimates never implication that ridge regression is incapable of performing variable selection. The rational of this method, variables with less contribution shrunk towards zero but it incorporates all the variables in the model (Farebrother, 1976) ; (Theobald, 1974). This is useful when all variables need to be incorporated in the model. Thus, the need to compare these restrictions and see if it will perform better when the outcome is a function of many predictors, all with coefficients of roughly equal size (James et al. 2014).

Model Development

In RR, we consider a sample of n observations,i.e, a training set data of $(x_1, y_1), ..., (X_n, Y_n)$, each consisting p variables and a single response(outcome). The mostly used estimation method is the least squares method which minimizes the residual sum of squares(RSS)

$$RSS = \sum_{i=1}^{n} (yi - \beta_0 - \sum_{j=i}^{p} \beta_j x_{ij})^2 \tag{3.1}$$

$$L_{RR}(\beta) = \sum_{i=1}^{n}(yi - \beta_o - \sum_{j=1}^{p}\beta_j x_i)^2 + \lambda \sum_{j=1}^{p}\beta_j^2 \tag{3.2}$$

which equals

$$RSS + \lambda \sum_{j=1}^{p}\beta_j^2 \tag{3.3}$$

Just like least squares, in the cost function is altered by adding the penalty square of magnitude of the coefficients

$$C = SSE + \lambda \sum_{j=0}^{p}\hat{\beta}_j^2 = \sum_{i=1}^{n}(Yi - \hat{y}_i)^2 + \lambda \sum_{j=0}^{p}\hat{\beta}_j^{\,2} = \sum_{i=1}^{n}\left[Y_i - \sum_{j=0}^{p}\hat{\beta}_j X_{ij}\right]^2 + \lambda \sum_{j=0}^{p}\hat{\beta}_j^{\,2} \tag{3.4}$$

Which is also expressed as

$$C = (y - X\hat{\beta})'(y - X\hat{\beta}) + \lambda\hat{\beta}'\hat{\beta} \tag{3.5}$$

Differentiating with respect to $\hat{\beta}$ we get

$$-2X'y + 2X'X\hat{\beta} + 2\lambda\hat{\beta} = 0. \tag{3.6}$$

then the Ridge regression estimator of β becomes

$$\hat{\beta}_{Ridge} = (X'X + \lambda I)^{-1}X'Y \tag{3.7}$$

Where $\lambda \geq 0$ is the tuning parameter controlling the regularization. The greater the size of λ the high the shrinkage. Whereas $\lambda\sum_{j=1}^{p}\beta_j^2$ is the penalty term which shrinks towards zero. That is, as $\lambda \to 0, \hat{\beta}^{RR} \to \hat{\beta}^{RSS}$ The rational of this approach is that it improves prediction error by shrinking large regression coefficients in order to reduce over-fitting, but it does not perform covariate selection and therefore does not help to make the model more interpretable. It is important to select a good value of λ hence we employ cross validation.

3.2.2 Lasso Regression (LR)

The Least Absolute Shrinkage and Selection Operator (Lasso) regression was first introduced by Tibshirani in 1996, is a regression technique that performs both regularization through penalizing and shrinking parameter estimates. It has the capability of doing variable selection by shrinking parameter estimates to zero exactly (Tibshirani, 1996).

This technique uses ι_1 penalty shrinks estimates to exactly zero when the tuning parameter is large, thus performing variable selection,achieving sparsity with a clear and interpretable model is the advantage of this model despite the loss of variables.

Model development

Just like the Ridge regression, LR is defined by,

$$L_{LR}(\beta) = \sum_{i=1}^{n}(yi - \beta_0 \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} |\beta_j| = RSS + \lambda \sum_{j=1}^{p} |\beta_j| \tag{3.8}$$

Let

$$(\hat{\beta}) = (\hat{\beta}_1 \ldots, \hat{\beta}_p)^T \tag{3.9}$$

then the lasso estimates $(\hat{\alpha}, \hat{\beta})$ is defined as

$$(\hat{\alpha}, \hat{\beta}) = argmin \sum_{j=1}^{N} (Yi - \alpha - \sum_{j} \beta_j x_{ij})^2 \tag{3.10}$$

subject to $\sum_j |\beta_j| \leq t$

where $t \geq 0$ is the tuning parameter.

The only change is that β_j^2 in ridge is replaced with $|\beta_j|$ in lasso, the ι_1 penalty . λ which is the tuning parameter which controls the amount of shrinkage. In this case, coefficients are shrunk to zero. This technique still employs cross-validation to come up with the best value of λ our tuning parameter.

3.2.3 Relaxed/Flexible Lasso

The relaxed Lasso was introduced by Meinshausen as a solution that extended, and addressed shortcomings of the Lasso (Meinshausen, 2007). Its target was to improve the bias of Lasso coefficients. To achieve the relaxed Lasso, a 2-step procedure is outlined,

Its estimator defined as,

$$\hat{\beta} = argmin \left[\frac{1}{N} \sum_{i=1}^{N} (y_i - x_i b)^2 + \lambda_{init} \lambda_{relaxed} \sum_{j=1}^{p} \frac{|b_j|}{|\hat{\beta_{j,init}}|} \right] \tag{3.11}$$

and the two steps are;

1. Fit the model with the Lasso penalty, and identify the non-zero coefficients.

2. Refit the same model without the Lasso penalty, while using only the covariates that correspond to the non-zero coefficients identified in 1 above.

The fact that Lasso estimates are completely shrunk, it is biased to zero. Therefore, by refitting the same model with the covariates selected from the Lasso model, without penalizing the coefficients, we are in other words "relaxing" (removing the penalty) the results we would obtain are simply the Lasso model.

VT OMNES VNVM SINT

3.2.4 Elastic -Net Regression (ENR)

This method was introduced by Zou and Hastie as a regularization approach which combines both ridge regression and lasso regression to select the appropriate influential variables and produce an accurate prediction model Zou and Hastie (2005). ENR aims at minimizing the loss function as defined,

$$J(\beta, \lambda_1, \lambda_2) = \sum_{j=i}^{p} \left[\lambda_1 |\beta_j| + \lambda_2 \beta_j^2 \right] \tag{3.12}$$

Consider $(x_i, y_i), i = 1, 2, 3, \ldots, N$, of i.i.d random vectors where $(x_{i1}, x_{i2}, ..., x_{ip})$ is the row vector p observations for the $i_t h$ sample unit and y_i is the response vector.

We assume the response is centered and predictors are standardized. That is, $\sum_{i=1}^{N} y_i, \sum_{j=1}^{N} x_i j = 0$ and $\sum_{j=1}^{N} x_{ij}^2 = 1$, for $j = 1, 2, \ldots, p$.

For any fixed non-negative λ_1 and λ_2, then the elastic-net estimator is defined as

$$\hat{\beta}_{enet} = (1 + \lambda_2) = argmin \left[\frac{1}{N} \sum_{i=1}^{N} (y_i - x_i b)^2 + \lambda_1 \sum_{j=1}^{p} |b_j| + \lambda_2 \sum_{j=1}^{p} b_j^2 \right] \tag{3.13}$$

In this case α is the controlling connecting parameter or mixing parameter between Ridge ($\alpha = 1$) and Lasso ($\alpha = 0$). This gives us two tuning parameters, that is λ and α. When α is set as 1 the model turns to ridge and so if set to 0 it produces Lasso model. It is therefore important to set α between $0 \to 1$. In other words, ENR is a linear combination of the ι_1 and ι_2 penalties. This technique just like a big fishing net, it's flexible which is able to contain many fish. Therefore, it is expected to give a better prediction and with optimal variable selection.

3.2.5 Data Simulation and Design

To test the performance and validity of the proposed approaches, we present experimental results on simulated data to mimic our real data from our ECMC. As stated, our simulated data will assume normal distribution. Normal distribution because using our real data we were able to overlay a normal curve. Under this assumption, we simulate 100 datasets to mimic our real data. Thereafter, fit our models as outlined in below simulation design.

a) n > p

n = 60, p = 20

n = 36, p = 20

n = 24, p = 20

b) n = p

n = 20, p = 20

c) n < p

n = 18, p = 20

n = 16, p = 20

In addition, we will then evaluate the RMSE, MAE and R^2 values which are our metrics of evaluation. Therefore, we will be able to compare the empirical results whenever parameters are adjusted and theoretically be informed on which model works and in which setting. Hence, we assess the prediction accuracy, overall model interpret-ability and robustness.

Set up.

For each of the cases, we fit the following models and assess their performance.

1. Logistic regression model with lasso penalty
2. Logistic regression model with ridge penalty
3. Logistic regression model with elastic net penalty

These models were built with R software using the glmnet package excluding the elastic net, which requires an additional package, the caret package, due to the existence of two tuning parameters. In addition, we apply cross validation approach to choose an appropriate value for our tuning parameters i.e. λ, and then fit logistic regression models to our dataset with the different regularization methods discussed under the different varying n, p conditions shown above, and present thereafter results and analysis.

Chapter 4

Results and Analysis

In this chapter, we present performance comparison of the three prediction models under the adjusted values of p and n. These are three different scenarios testing the performance of three different models. Our measures of performance are RMSE, MAE and R^2 which well be looking at for each model. First, we evaluate model performance on simulated data, select the best performing model, then use it to fit our sales data.

4.1 n > p Scenario

In these scenario we did three adjustments on our p,n values. For (n=60,p=20), the Elastic net model performed better compared to ridge and lasso based RMSE and MAE. However, performance based on R^2, the lasso model did the best as shown in Table 4.1.

Table 4.1: Comparison of regression models performance when n=60 and p=20

Model	RMSE	MAE	R^2
Elastic net	39,340,379	32,216,733	0.08977566
Lasso	39,774,772	32,569,428	0.09239257
Ridge	49,354,421	40,109,938	0.08930728

Under (n=36, p=20) adjustment, elastic net performed the best based on the values of RMSE and MAE while based on R^2 Lasso did extremely well as shown in Table 4.2.

Table 4.2: Comparison of regression models performance when n = 36, p = 20

Model	RMSE	MAE	R^2
Elastic net	40,027,009	33,246,145	0.2009075
Lasso	40,061,285	32,956,201	0.1559531
Ridge	56,092,075	46,269,033	0.1763819

Still in n > p scenario, we adjusted (n=24 ,p=20). On this this particular adjustments, the Elastic net model proved to be the best based on the values of the three metrics as shown in Table 4.3.

Table 4.3: Comparison of regression models performance when , n = 24, p = 20

Model	RMSE	MAE	R^2
Elasticnet	39,542,737	33,582,773	0.2974244
Lasso	42,814,170	35,899,888	0.2529814
Ridge	61,429,503	50,669,563	0.2558258

4.2 n = p Scenario

This section presents the results of (n = p) setting. Out this setting. In this setting Elastic net still outdid the rest in terms of RMSE, MAE and R^2 as seen in Table 4.4.

Table 4.4: Comparison of regression models performance when the number of obeservations is equal to the number of predictors n = 20, p =20

Model	RMSE	MAE	R^2
Elasticnet	36,420,875	30,918,412	0.3719545
Lasso	40,215,506	34,508,193	0.3537908
Ridge	54,568,443	46,536,368	0.3562105

4.3 n < p Scenario

This section presents thee last scenario of our study,(n < p). Under this scenario, we adjusted our p and n values twice. In the first one (n = 18, p = 20), Elastic net performed better based on RMSE and MAE. Contrary, based of R^2 the ridge model did extremely good than all as shown in Table 4.5.

Table 4.5: Comparison of regression models performance when n = 18, p = 20

Model	RMSE	MAE	R^2
Elasticnet	42,595,673	35,614,376	0.2839778
Lasso	43,482,498	36,376,549	0.3626423
Ridge	62,424,737	52,362,093	0.3738885

The second and final adjustment under this scenario (n = 16, p = 20), The elastic net gave the desired model based that elastic net is the desired model. However based on R^2 Ridge gives the best model . However, based on R^2 Ridge was the best model as shown in Table 4.6 .

Table 4.6: Comparison of regression models performance when n = 16, p = 20

Model	RMSE	MAE	R^2
Elasticnet	39,676,031	34,136,704	0.4606201
Lasso	39,974,337	34,196,528	0.4928461
Ridge	55,793,738	48,202,767	0.5311499

4.4 Regression Model On Sampled Sales Data

Based on above results of models on simulated data. We choose elastic net as our desired model and fit it on our real data. The results of elastic net on our real data gave below out puts based on RMSE, MAE and R^2 as shown in Table 4.7.

The obtained values of $\alpha = 0.3$ and $\lambda = 795669.6$

Table 4.7: Elastic Net Regression model Performance Results on sampled sales data

RMSE.net	MAE.net	R^2.net
2545127	2006233	0.99951

Elastic net regression model plot on variable selection is shown in Figure 4.1.

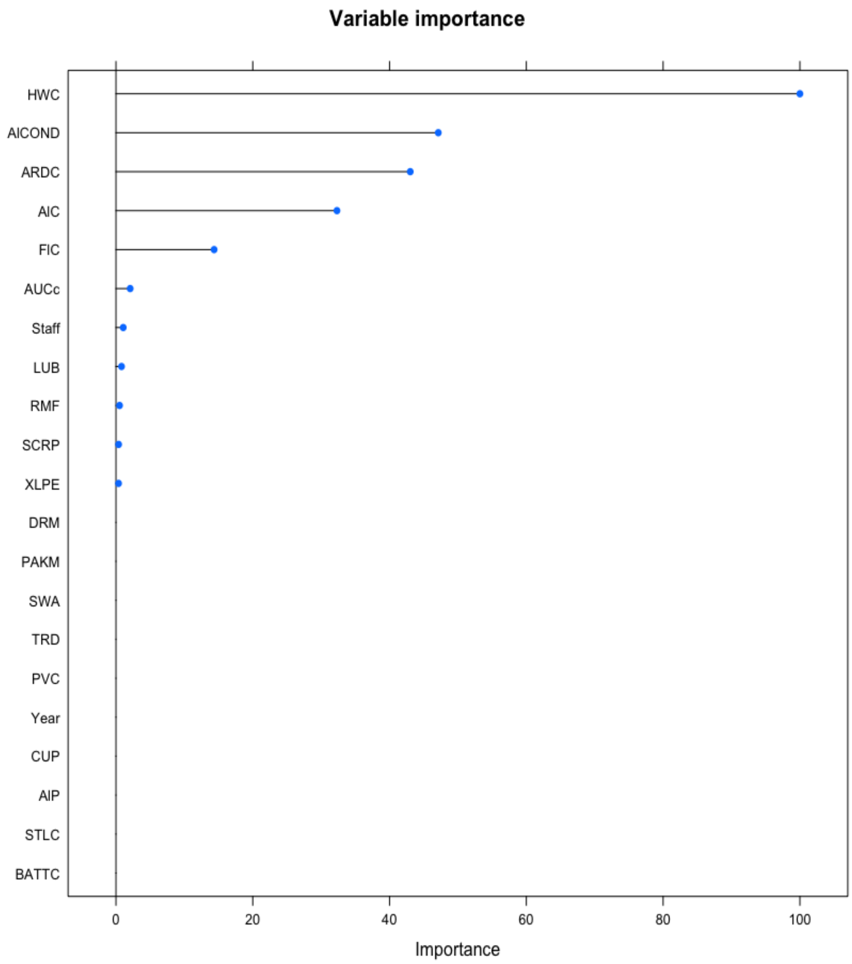

Figure 4.1: Elastic Net Regression Model Variable Plot.

VT OMNES VNVM SINT

Chapter 5

Discussion and Conclusion

This chapter presents a brief discussion of the analyzed results obtained in chapter 4 above. These include, the simulated data results, model fit under different scenarios and the fit model on our sampled data. We compared performance of three regularized approaches systematically, under the selected varying values of p and n. To start, we present an overview of our evaluating metrics on how they are used and present their obtained values under the scenarios presented. Afterwards, analyze the performance of the models based on these metrics

5.1 Evaluation metrics

In this section we look at our three evaluation metrics the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and R-SquaredR^2. RMSE is the square root of Mean Squared error (MSE), it indicates the absolute fit of the model to a data i.e. shows how close the observed data points are to predicted values by the model. In other words, it measures how accurately a model predicts the response. As a result, this metric is widely used to evaluate the performance of regression models. The lower the value of RMSE the better the fit .

Mean absolute error represents the average of the absolute difference between the predicted and the actual values in the dataset. It measures the average of the residuals in the dataset. R-squared or the coefficient of determination represents the proportion of the variance in the response variable or the output of a regression model i.e. the proportion of variance in the dependent variable which is explained by the linear regression model. Its values are neither small nor large i.e. it is a scale-free score its value is less than one. R^2 explains how well, the independent variables in the linear regression model explain the variability in the dependent variable. R-Squared increases with the addition of the independent variables. It puts into account the number of predictor variables, and used to determine the number of independent variables in a model. Both RMSE and R^2 explain how well a model fits a dataset. While the RMSE tells how well a regression model can predict, the R^2 inform us on how well the predictor variables can explain the variation in the response variable. RMSE is more preferred for accuracy comparison among different prediction to R-Squared, however RMSE is sensible to the inclusion of additional variables in a model, even if those variables don't have high significant contribution in explaining the outcome Kassambara (2018).

5.2 Model performance

We start with discussing our simulation results. In our simulation section, we explored different scenarios. In (n > p) setting, results presented in Table4.1 , our aim was to choose the model with the best prediction accuracy and interpret ability. From our first scenario, Elastic Net gave the best prediction with the lowest RMSE and MAE values followed by the Ridge and the the Lasso. Just like in the Zou and Zhang (2009) study, it was found that the elastic net often outperforms RR and the lasso in terms of model selection consistency and prediction accuracy.

Still under these scenario, we adjusted our n, p values and Elastic net did extraordinarily well on both accuracy and interpret-ability. It showed the lowest RMSE and MAE and a rather higher R^2 . Lasso followed up based on RMSE and MAE but Ridge over-performed Lasso based on R^2. Our second scenario was where the number of predictor is equal to the number of response variables. Under this setting, our results still gave Elastic net as the desired model.

The final scenario and the highlight of this study (high dimensionality), p > n setting. Under the first adjustments, Elastic net gave best prediction based of RMSE and MAE but Ridge was superior based on R^2. The final adjustment Elastic net gave the best prediction model with the RMSE and MAE. However, even though elastic net possesses the oracle property and is more robust to multicollinearity, and hence would be expected to have high predictive accuracy, it had lower accuracy than the RR and Ridge when evaluating using the metric R^2.

Based on these results, we chose Elastic net which is a combination of Ridge and Lasso as our best prediction model. Just like in Khashei and Bijari (2010), Elastic Net performed extremely well compared to the aggregated approaches separately. Just like reviewed studies on prediction using hybrid models Zhang (2003), Elastic net which combines both restrictions of the Ridge and Lasso demonstrated superior performance. It dealt well with changing market dynamics. Therefore, Elastic net became our desired prediction model and therefore, we used it to fit our real sales data.The results on on Elastic net regression model on real data was as presented on table above. Additionally, we tuned our variables of importance on our model and got the displayed important variable plot shown previously.

5.3 Conclusion and Recommendations

In this research, we presented a complete outcome of comparison of sales prediction. The methodology is particularly useful for manufacturers , i.e ECMCs whereby, with accurate sales prediction, finance department is able to project costs, profits and capital budgeting. The sales department will have a good estimates on sales volumes of each product and hence able to plan for sales labour force. Management team(COO) is able to do production plan, raw material acquisition and logistics correctly without incurring undesired expenses.

However, there are different inconsistencies with our results. For instance, elastic net regression model has performed excellent on model prediction with the minimum error across all scenarios i.e it produced the lowest number of RMSE and MAE. This proves its flexibility capability to handle unavoidable changes in our today's competitive business dynamics. Despite having the most accurate prediction, Ridge and Lasso regression models at some instances had superior performance. For example, in the scenario (n=18,p=20) and (n=60,p=20) Lasso and Ridge regression did better compared to Elastic net regression when evaluated using R^2. Despite ENR model performing excellent based on RMSE and MAE its still failed under different circumstances using R-squared as mentioned in above scenarios.

Our expectations was that the best model (one with the smallest RMSE and MAE) would have the highest R^2. However, due to the sensible inclusion of additional variables in a model by RMSE, the R^2 Zou and Zhang (2009) is expected to be highest when RMSE is lowest which varied. our recommendations for any further studies on this topic is to explain the cited variation.

VT OMNES VNVM SINT

References

Aburto, L. and Weber, R. (2007). Improved supply chain management based on hybrid demand forecasts. *Applied Soft Computing*, 7(1):136–144.

Adya, M. and Collopy, F. (1998). How effective are neural networks at forecasting and prediction? a review and evaluation. *Journal of forecasting*, 17(5-6):481–495.

Ainscough, T. L. and Aronson, J. E. (1999). An empirical investigation and comparison of neural networks and regression for scanner data analysis. *Journal of Retailing and Consumer Services*, 6(4):205–217.

Aslanargun, A., Mammadov, M., Yazici, B., and Yolacan, S. (2007). Comparison of arima, neural networks and hybrid models in time series: tourist arrival forecasting. *Journal of Statistical Computation and Simulation*, 77(1):29–53.

Bühlmann, P. and Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Callen, J. L., Kwan, C. C., Yip, P. C., and Yuan, Y. (1996). Neural network forecasting of quarterly accounting earnings. *International Journal of Forecasting*, 12(4):475–482.

Chang, P.-C. and Wang, Y.-W. (2006). Fuzzy delphi and back-propagation model for sales forecasting in pcb industry. *Expert systems with applications*, 30(4):715–726.

Chen, F. and Ou, T. (2008). A neural-network-based forecasting method for ordering perishable food in convenience stores. In *2008 fourth international conference on natural computation*, volume 2, pages 250–254. IEEE.

Chu, C.-W. and Zhang, G. P. (2003). A comparative study of linear and nonlinear models for aggregate retail sales forecasting. *International Journal of production economics*, 86(3):217–231.

Church, K. B. and Curram, S. P. (1996). Forecasting consumers' expenditure: A comparison between econometric and neural network models. *International journal of forecasting*, 12(2):255–267.

Crone, S. F. (2003). Prediction of white noise time series using artificial neural networks and asymmetric cost functions. In *Proceedings of the International Joint Conference on Neural Networks, 2003.*, volume 4, pages 2460–2465. IEEE.

Díaz-Robles, L. A., Ortega, J. C., Fu, J. S., Reed, G. D., Chow, J. C., Watson, J. G., and Moncada-Herrera, J. A. (2008). A hybrid arima and artificial neural networks model to forecast particulate matter in urban areas: The case of temuco, chile. *Atmospheric Environment*, 42(35):8331–8340.

Doganis, P., Alexandridis, A., Patrinos, P., and Sarimveis, H. (2006). Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. *Journal of Food Engineering*, 75(2):196–204.

VT OMNES VNVM SINT

Farebrother, R. W. (1976). Further results on the mean square error of ridge regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 248–250.

Foster, W., Collopy, F., and Ungar, L. (1992). Neural network forecasting of short, noisy time series. *Computers & chemical engineering*, 16(4):293–297.

Frank, C., Garg, A., Sztandera, L., and Raheja, A. (2003). Forecasting women's apparel sales using mathematical modeling. *International Journal of Clothing Science and Technology*.

Greene, W. H. (1981). On the asymptotic bias of the ordinary least squares estimator of the tobit model. *Econometrica: Journal of the Econometric Society*, pages 505–513.

Gupta, A., Maranas, C. D., and McDonald, C. M. (2000). Mid-term supply chain planning under demand uncertainty: customer demand satisfaction and inventory management. *Computers & Chemical Engineering*, 24(12):2613–2621.

Hanssens, D. M., Parsons, L. J., and Schultz, R. L. (2003). *Market response models: Econometric and time series analysis*, volume 2. Springer Science & Business Media.

Heravi, S., Osborn, D. R., and Birchenhall, C. (2004). Linear versus neural network forecasts for european industrial production series. *International Journal of Forecasting*, 20(3):435–446.

Hill, T., Marquez, L., O'Connor, M., and Remus, W. (1994). Artificial neural network models for forecasting and decision making. *International journal of forecasting*, 10(1):5–15.

Jain, A. and Kumar, A. M. (2007). Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing*, 7(2):585–592.

Kassambara, A. (2018). *Machine learning essentials: Practical guide in R*. Sthda.

Khashei, M. and Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with applications*, 37(1):479–489.

Kubat, M. (1999). Neural networks: a comprehensive foundation by simon haykin, macmillan, 1994, isbn 0-02-352781-7. *The Knowledge Engineering Review*, 13(4):409–412.

Kuo, R. and Cohen, P. (1998). Intelligent tool wear estimation system through artificial neural networks and fuzzy modeling. *Artificial Intelligence in Engineering*, 12(3):229–242.

Kuo, R. J., Wu, P., and Wang, C. (2002). An intelligent sales forecasting system through integration of artificial neural networks and fuzzy neural networks with fuzzy weight elimination. *Neural networks*, 15(7):909–925.

Meinshausen, N. (2007). Relaxed lasso. *Computational Statistics & Data Analysis*, 52(1):374–393.

Mentzer, J. T. and Bienstock, C. C. (1998a). *Sales forecasting management: understanding the techniques, systems and management of the sales forecasting process*. SAGE Publications, Incorporated.

Mentzer, J. T. and Bienstock, C. C. (1998b). The seven principles of sales-forecasting systems. *Supply Chain Management Review*, 34(4):76–83.

Nelson, M., Hill, T., Remus, B., and O'Connor, M. (1994). Can neural networks applied to time series forecasting learn seasonal patterns: an empirical investigation. In *1994 Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences*.

Ntungo, C. and Boyd, M. (1998). Commodity futures trading performance using neural network models versus arima models. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 18(8):965–983.

Ogutu, J. O., Schulz-Streeck, T., and Piepho, H.-P. (2012). Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. In *BMC proceedings*, volume 6, pages 1–6. Springer.

Paliwal, M. and Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert systems with applications*, 36(1):2–17.

Qi, M. (2001). Predicting us recessions with leading indicators via neural network models. *International journal of forecasting*, 17(3):383–401.

Ranstam, J. and Cook, J. (2018). Lasso regression. *Journal of British Surgery*, 105(10):1348–1348.

Reddy, S. K., Holak, S. L., and Bhat, S. (1994). To extend or not to extend: Success determinants of line extensions. *Journal of marketing research*, 31(2):243–262.

Song, Q. and Esogbue, A. O. (2006). A new algorithm for automated box-jenkins arma time series modeling using residual autocorrelation/partial autocorrelation functions. *Industrial Engineering and Management Systems*, 5(2):116–125.

Sun, Z.-L., Choi, T.-M., Au, K.-F., and Yu, Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. *Decision Support Systems*, 46(1):411–419.

Tang, Z., De Almeida, C., and Fishwick, P. A. (1991). Time series forecasting using neural networks vs. box-jenkins methodology. *Simulation*, 57(5):303–310.

Theobald, C. M. (1974). Generalizations of mean square error applied to ridge regression. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(1):103–106.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Wang, L., Zou, H., Su, J., Li, L., and Chaudhry, S. (2013). An arima-ann hybrid model for time series forecasting. *Systems Research and Behavioral Science*, 30(3):244–259.

Yip, D. H., Hines, E. L., and Yu, W. W. (1997). Application of artificial neural networks in sales forecasting. In *Proceedings of International Conference on Neural Networks (ICNN'97)*, volume 4, pages 2121–2124. IEEE.

Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1):35–62.

Zhang, G. P. (2003). Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175.

Zhang, G. P. (2004). *Neural networks in business forecasting*. IGI global.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320.

Zou, H. and Zhang, H. H. (2009). son the adaptive elastic-net with a diverging number of parameter. *Annals of statistics*, 37(4):1733.

Appendix A

R code

The code used in Chapter 3 and Chapter 4.

A.1 Data distribution

```
#Load data and view

salesdata <- read.csv(’fulldata.csv’)



# X-axis grid
options(scipen = 1000)
x <- salesdata$VOLUME
 # X-axis grid
x2 <- seq(min(x), max(x), length = 400)
# Normal curve
fun <- dnorm(x2, mean = mean(x), sd = sd(x))

# Histogram
hist(x, prob = TRUE, col = "white",
     ylim = c(0, 0.000000015),
     main = "Histogram of sales with normal curve")
lines(x2, fun, col = 2, lwd = 2)
```

A.2 Data Simulations

```
# Script to generate simulated data

# user function to generate 1 dataset
modata <-
  function(ns, p){
    # generate the dependent variable
    VOLUME <- rnorm(n = ns, mean = 138904259 , sd = 38580046)
    # generate independent variables
    Staff <- round(rnorm(n = ns, mean = 20, sd = 5.86))
    RMF <- rnorm(n = ns, mean =170361.0  , sd = 257607.4)
    AUCc <- rnorm(n = ns, mean =  405081 , sd = 1114779)
    HWC <- rnorm(n = ns, mean = 77517659 , sd = 21926460)
    ARDC <- rnorm(n = ns, mean = 27390442 , sd = 11089646)
    FlC <- rnorm(n = ns, mean = 5738664  , sd = 3218340)
    AlC <- rnorm(n = ns, mean = 12459669 , sd = 13565400)
    AlCOND <- rnorm(n = ns, mean = 14936867  , sd = 16733031)
    STLC <- rnorm(n = ns, mean = 29281, sd = 36439.86)
    SCRP <- rnorm(n = ns, mean = 174880, sd = 246958.6)
    DRM <- rnorm(n = ns, mean = 228.0  , sd = 212.05)
    TRD <- rnorm(n = ns, mean = 99791  , sd = 136066.8)
    BATTC <- rnorm(n = ns, mean = 500263 , sd = 642952.9)
    AlP <- rnorm(n = ns, mean = 14588855    , sd = 11045763)
    CUP <- rnorm(n = ns, mean = 13311580   , sd = 13370810)
    SWA <- rnorm(n = ns, mean = 26017821, sd = 5398313)
    XLPE <- rnorm(n = ns, mean = 14496676   , sd = 4019579)
    PVC <- rnorm(n = ns, mean = 16870734 , sd = 11191620)
    LUB <- rnorm(n = ns, mean = 710932   , sd = 171378.3)
    PAKM <- rnorm(n = ns, mean = 627604 , sd = 493629.2)
```

```
    # combine to data frame
    df <- cbind(VOLUME, Staff, RMF, AUCc, HWC, ARDC, FlC,
                AlC, AlCOND, STLC, SCRP, DRM, TRD,
                BATTC, AlP, CUP, SWA, XLPE, PVC,
                LUB, PAKM)
    # Output the data
    df[,1:(p+1)]
  }

# Simulation multiplier
# generate many datasets in specific subfolders

multiplier <-
  function(n = 50, ns2 = 60, p2 = 20, folder = "ngreaterp",
           sub = "sim_60_20"){
    # Set empty list
    daf = list()

    for(i in 1:n){
      # create data
      df <- modata(ns = ns2,  p = p2 )
      daf[[i]] <- df}
    # Write out list as csv files
    for(i in 1:length(daf)){
      write.csv(data.frame(daf[[i]]),
                file = paste0("simulation/", folder, "/", sub, "/",ns2,"_",i, '.csv
    }

  }
```

```
# Generate instances where n > p
# 5 years (n = 60, p =20)
# set.seed
set.seed(432)
# get data
multiplier(n = 100, ns2 = 60, p2 = 20, folder = "ngreaterp",
           sub = "sim_60_20")

# 3 years (n = 36, p =20)
# set.seed
set.seed(432)
# get data
multiplier(n = 100, ns2 = 36, p2 = 20, folder = "ngreaterp",
           sub = "sim_36_20")

# 1.5 years (n = 24, p =20)
# set.seed
set.seed(432)
# get data
multiplier(n = 100, ns2 = 24, p2 = 20, folder = "ngreaterp",
           sub = "sim_24_20")

###

# Generate instances where n = p
# 1 year (n = 20, p =20)
# set.seed
set.seed(432)
# get data
```

```
multiplier(n = 100, ns2 = 20, p2 = 20, folder = "nequalp",
           sub = "sim_20_20")


####

# Generate instances where n < p
# 10 months years (n = 18, p =20)
# set.seed
set.seed(432)
# get data
multiplier(n = 100, ns2 = 18, p2 = 20, folder = "nlessp",
           sub = "sim_18_20")

# 6 months (n = 16, p =20)
# set.seed
set.seed(432)
# get data
multiplier(n = 100, ns2 = 16, p2 = 20, folder = "nlessp",
           sub = "sim_16_20")
```

A.3 Simulation of models

Script to fit model and store results

```
# load libraries

library(caret)
```

```
library(elasticnet)
library(glmnet)

# Run ridge regression
ridger <-
  function(x){
    # create training and testing datasets
    parter <- round(length(x$VOLUME) * 0.80)
    traindata <- x[1:parter,-1]
    testdata <- x[(parter+1):nrow(x),-1]

    # 5 fold cross validation

    ctrl <- trainControl(
      method = "cv",
      number = 5,
    )

    # fit model
    rmodel <- train(
      VOLUME ~ .,
      data = traindata,
      method = 'ridge',
      preProcess = c("center", "scale"),
      trControl = ctrl

    )

    # make predictions on test data
    predictions = predict(rmodel, newdata = testdata)
```

VT OMNES VNVM SINT

```
    # get metrics
    met <- cbind(n = nrow(x),
                 p = ncol(x)-2,
                 RMSE = RMSE(pred = predictions,
                             obs = testdata$VOLUME, na.rm = F),
                 MAE = MAE(pred = predictions,
                           obs = testdata$VOLUME, na.rm = F),
                 R2 = R2(pred = predictions,
                         obs = testdata$VOLUME, na.rm = F))
    met <- data.frame(met)
    # add model name
    met$model <- "ridge"
    # result
    met
  }


# Run lasso regression

mlasso <-
  function(x){
    # create training and testing datasets
    parter <- round(length(x$VOLUME) * 0.80)
    traindata <- x[1:parter,-1]
    testdata <- x[(parter+1):nrow(x),-1]

    # 10 fold cross validation
    ctrl <- trainControl(
      method = "cv",
```

VT OMNES VNVM SINT

```
  number = 5,
)
rownames(traindata) <- NULL

# fit model
lmodel <- train(
  VOLUME ~ .,
  data = traindata,
  method = 'lasso',
  preProcess = c("center", "scale"),
  trControl = ctrl

)

# make predictions on test data
predictions = predict(lmodel, newdata = testdata)

# get metrics
met <- cbind(n = nrow(x),
             p = ncol(x)-2,
             RMSE = RMSE(pred = predictions,
                         obs = testdata$VOLUME, na.rm = F),
             MAE = MAE(pred = predictions,
                       obs = testdata$VOLUME, na.rm = F),
             R2 = R2(pred = predictions,
                     obs = testdata$VOLUME, na.rm = F))
met <- data.frame(met)
# add model name
met$model <- "lasso"
# result
```

```
    met
  }

# Run elastic net regression

elasto <-
  function(x){
    # create training and testing datasets
    parter <- round(length(x$VOLUME) * 0.80)
    traindata <- x[1:parter,-1]
    testdata <- x[(parter+1):nrow(x),-1]

    # 10 fold cross validation
    ctrl <- trainControl(
      method = "cv",
      number = 5

    )

    # fit model
    rmodel <- train(
      VOLUME ~ .,
      data = traindata,
      method = 'glmnet',
      preProcess = c("center", "scale"),
      trControl = ctrl,
      tuneLength = 10

    )
```

VT OMNES VNVM SINT

```
    # make predictions on test data
    predictions = predict(rmodel, newdata = testdata)

    # get metrics
    met <- cbind(n = nrow(x),
                 p = ncol(x)-2,
                 RMSE = RMSE(pred = predictions,
                             obs = testdata$VOLUME, na.rm = F),
                 MAE = MAE(pred = predictions,
                           obs = testdata$VOLUME, na.rm = F),
                 R2 = R2(pred = predictions,
                           obs = testdata$VOLUME, na.rm = F))
    met <- data.frame(met)
    # add model name
    met$model <- "elasticnet"
    # result
    met
  }

# Function to run all models
modeller <-
  function(x){
    # model results
    rd_result <- ridger(x)
    las_result <- mlasso(x)
    ela_result <- elasto(x)



    # Combine results
    result <- rbind(rd_result,
```

```
                    las_result,
                    ela_result)
    result
  }
```

A.4 Modelling on Simulated Data

```
n the simulations
# source modeller script
source('models_script.R')
# load library
library(purrr)
library(dplyr)

# Models where n > p

# Set seed
set.seed(435)

# Results for 60 months
sixty20 <- list.files(path = "./simulation/ngreaterp/sim_60_20",
                             # Identify all csv files in folder
                             pattern = "*.csv", full.names = TRUE) %>%
  lapply(read.csv) %>%             # Store all files in list
  map(modeller)
sixty20_df <- do.call(rbind, sixty20)        # Combine data sets into one data set
```

```
# write.csv
write.csv(sixty20_df, "./model_results/sixty_20_metrics.csv")



# Results for 36 months
thirty620 <- list.files(path = "./simulation/ngreaterp/sim_36_20",
                        # Identify all csv files in folder
                        pattern = "*.csv", full.names = TRUE) %>%
  lapply(read.csv) %>%              # Store all files in list
  map(modeller)
thirty620_df <- do.call(rbind, thirty620)        # Combine data sets into one data
# write.csv
write.csv(thirty620_df, "./model_results/thirty6_20_metrics.csv")

# Results for 24 months
twofour20 <- list.files(path = "./simulation/ngreaterp/sim_24_20",
                          # Identify all csv files in folder
                          pattern = "*.csv", full.names = TRUE) %>%
  lapply(read.csv) %>%              # Store all files in list
  map(modeller)
twofour20_df <- do.call(rbind, twofour20)        # Combine data sets into one data
# write.csv
write.csv(twofour20_df, "./model_results/twofour_20_metrics.csv")



# Models where n = p
################



# Results for 20 months
```

```
twenty20 <- list.files(path = "./simulation/nequalp/sim_20_20",
                      # Identify all csv files in folder
                      pattern = "*.csv", full.names = TRUE) %>%
  lapply(read.csv) %>%            # Store all files in list
  map(modeller)
twenty20_df <- do.call(rbind, twenty20)       # Combine data sets into one data se
# write.csv
write.csv(twenty20_df, "./model_results/twenty_20_metrics.csv")

# Models where n < p

# Results for 18 months
eighteen20 <- list.files(path = "./simulation/nlessp/sim_18_20",
                       # Identify all csv files in folder
                       pattern = "*.csv", full.names = TRUE) %>%
  lapply(read.csv) %>%            # Store all files in list
  map(modeller)
eighteen20_df <- do.call(rbind, eighteen20)       # Combine data sets into one dat
# write.csv
write.csv(eighteen20_df, "./model_results/eighteen_20_metrics.csv")

# Results for 16 months
sixteen20 <- list.files(path = "./simulation/nlessp/sim_16_20",
                    # Identify all csv files in folder
                    pattern = "*.csv", full.names = TRUE) %>%
  lapply(read.csv) %>%            # Store all files in list
  map(modeller)
sixteen20_df <- do.call(rbind, sixteen20)       # Combine data sets into one data
# write.csv
write.csv(sixteen20_df, "./model_results/sixteen_20_metrics.csv")
```

A.5 Modelling on Sampled Data

```
# Analysis


## ------------------------------------------------------------------------------------------------
# load library
library(flextable)
# read data
ngrp_60 <- read.csv("model_results/sixty_20_metrics.csv")
library(dplyr)
ngrp_60 %>%
  group_by(model) %>%
  summarise(mean_RMSE = mean(RMSE),
            mean_MAE = mean(MAE),
            mean_R2 = mean(R2)) %>%
  regulartable() %>% autofit()


## ------------------------------------------------------------------------------------------------
# read data
ngrp_36 <- read.csv("model_results/thirty6_20_metrics.csv")
library(dplyr)
ngrp_36 %>%
  group_by(model) %>%
  summarise(mean_RMSE = mean(RMSE),
            mean_MAE = mean(MAE),
```

```
            mean_R2 = mean(R2))%>%
  regulartable()%>% autofit()


## -----------------------------------------------------------------------------------------
# read data
ngrp_24 <- read.csv("model_results/twofour_20_metrics.csv")
library(dplyr)
ngrp_24 %>%
  group_by(model) %>%
  summarise(mean_RMSE = mean(RMSE),
            mean_MAE = mean(MAE),
            mean_R2 = mean(R2, na.rm = T))%>%
  regulartable()%>% autofit()


## -----------------------------------------------------------------------------------------
# read data
nep_20 <- read.csv("model_results/twenty_20_metrics.csv")
library(dplyr)
nep_20 %>%
  group_by(model) %>%
  summarise(mean_RMSE = mean(RMSE),
            mean_MAE = mean(MAE),
            mean_R2 = mean(R2, na.rm = T))%>%
  regulartable()%>% autofit()


## -----------------------------------------------------------------------------------------
# read data
```

```
nlp_18 <- read.csv("model_results/eighteen_20_metrics.csv")
library(dplyr)
nlp_18 %>%
  group_by(model) %>%
  summarise(mean_RMSE = mean(RMSE),
            mean_MAE = mean(MAE),
            mean_R2 = mean(R2, na.rm = T))%>%
  regulartable()%>% autofit()


## ------------------------------------------------------------------------------------
# read data
nlp_16 <- read.csv("model_results/sixteen_20_metrics.csv")
library(dplyr)
nlp_16 %>%
  group_by(model) %>%
  summarise(mean_RMSE = mean(RMSE),
            mean_MAE = mean(MAE),
            mean_R2 = mean(R2, na.rm = T))%>%
  regulartable()%>% autofit()


## ------------------------------------------------------------------------------------
# read data
sales <- read.csv("salesdata.csv")


## ---- message=FALSE----------------------------------------------------------------
# load library
library(mice)
```

```
fd <- mice(sales,maxit=50,meth='cart',seed=500, m = 5)
fulldata <- complete(fd,1)
summary(fulldata)
# write full data
write.csv(fulldata,"fulldata.csv")


## ------------------------------------------------------------------------------------
library(caret)
set.seed(345)
trainIndex <- createDataPartition(fulldata$VOLUME, p = .8,
                                  list = FALSE,
                                  times = 1)
salesTrain <- fulldata[ trainIndex,-c(1,2)]
salesTest  <- fulldata[-trainIndex, -c(1,2)]


## ------------------------------------------------------------------------------------
model.net <- train(VOLUME ~., data = salesTrain, method = "glmnet",
                   trControl = trainControl("cv", number = 10),tuneLength = 10)

model.net$bestTune


## ---- eval=FALSE------------------------------------------------------------------------
## coef(model.net$finalModel, model.net$bestTune$lambda)
##
## x.test.net <- model.matrix(VOLUME ~., salesTest)[,-1]
## library(dplyr)
## predictions.net <- model.net %>% predict(x.test.net)
```

```
##
## data.frame(RMSE.net = RMSE(predictions.net, salesTest$VOLUME),
##            MAE.net = MAE(predictions.net, salesTest$VOLUME),
##            Rsquare.net = R2(predictions.net, salesTest$VOLUME))%>%
##   regulartable()%>% autofit()
##
```

```
## ------------------------------------------------------------------------
#source modeller script
source('models_script.R')
modeller(fulldata[,-1])%>%
  regulartable()%>% autofit()
```

```
## ------------------------------------------------------------------------
library(flextable)
#source modeller script
modeller(fulldata[1:20,-1])%>%
  regulartable()%>% autofit()
```

```
## ------------------------------------------------------------------------
# source modeller script
modeller(fulldata[1:16,-1])%>%
  regulartable()%>% autofit()
```

Appendix B

Ethical Approval

VT OMNES VNVM SINT

Strathmore UNIVERSITY

6th June 2022

Ms Muoki Monica
monica.muoki@strathmore.edu

Dear Ms Muoki,

RE: A Systematic Comparison of restrictions of Ridge, Lasso, Elastic Net and Relaxed Elastic approaches when fitting high dimensional data to predict sales.

This is to inform you that SU-IERC has reviewed and **approved** your above **SU Masters'** research proposal. Your application reference number is **SU-IERC1350/22.** The approval period is **6th June 2022 to 5th June 2023.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-IERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-IERC within 48 hours of notification
iv. Any changes, anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-IERC within 48 hours
v. Clearance for export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days upon completion of the study to SU-IERC.

Prior to commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

for: **Dr Ben Ngoye,**
Secretary; SU-IERC

Cc: Prof Fred Were,
Chairperson; SU-IERC

Appendix C

Similarity Report

Ouriginal

Document Information

Analyzed document	Thesis-Monica Muoki-MSc-SS.pdf (D139281468)
Submitted	2022-06-04T10:16:00.0000000
Submitted by	
Submitter email	Monica.Muoki@strathmore.edu
Similarity	2%
Analysis address	library.strath@analysis.urkund.com

Sources included in the report

W	URL: https://repository.tudelft.nl/islandora/object/uuid:699fd158-306f-40cc-a778-ea5ac5855b9c/datastream/OBJ/download Fetched: 2020-12-19T07:23:39.5700000	1
W	URL: https://www.hse.ru/data/2014/06/21/1310663809/ANNs.pdf Fetched: 2019-10-16T21:24:15.3530000	5
SA	**12212044.pdf** Document 12212044.pdf (D48397070)	3
W	URL: https://arxiv.org/pdf/2002.06384 Fetched: 2022-06-04T10:15:42.2170000	1
W	URL: https://dergipark.org.tr/en/pub/maruoneri/article/684425 Fetched: 2022-06-04T10:16:09.1400000	1
SA	**11050738.pdf** Document 11050738.pdf (D45167550)	1